ESCOLA DE GUERRA NAVAL

CC (T) ROSANA LEANDRO DE OLIVEIRA / C-Sup 2024

DEFESA CIBERNÉTICA:
O emprego de Sistemas de detecção de intrusão com apoio de aprendizado de máquina para otimizar a detecção de ataques

Rio de Janeiro
2024

CC (T) ROSANA LEANDRO DE OLIVEIRA/ C-Sup 2024

DEFESA CIBERNÉTICA:
O emprego de Sistemas de detecção de intrusão com apoio de aprendizado de máquina para otimizar a detecção de ataques

Monografia apresentada à Escola de Guerra Naval, como requisito parcial para a conclusão do Curso Superior.

Orientador: CC ROBERTO PIMENTA

Rio de Janeiro

Escola de Guerra Naval

2024

DECLARAÇÃO DA NÃO EXISTÊNCIA DE APROPRIAÇÃO INTELECTUAL IRREGULAR

Declaro que este trabalho acadêmico: a) corresponde ao resultado de investigação por mim desenvolvida, enquanto discente da Escola de Guerra Naval (EGN); b) é um trabalho original, ou seja, que não foi por mim anteriormente utilizado para fins acadêmicos ou quaisquer outros; c) é inédito, isto é, não foi ainda objeto de publicação; e d) é de minha integral e exclusiva autoria.

Declaro também que tenho ciência de que a utilização de ideias ou palavras de autoria de outrem, sem a devida identificação da fonte, e o uso de recursos de inteligência artificial no processo de escrita constituem grave falta ética, moral, legal e disciplinar. Ademais, assumo o compromisso de que este trabalho possa, a qualquer tempo, ser analisado para verificação de sua originalidade e ineditismo, por meio de ferramentas de detecção de similaridades ou por profissionais qualificados.

gov.br
Documento assinado digitalmente
ROSANA LEANDRO DE OLIVEIRA
Data: 22/11/2024 09:52:12-0300
Verifique em https://validar.iti.gov.br

DEDICATÓRIA

Dedico esta monografia à DEUS, por me sustentar com saúde e resiliência em mais esse desafio da minha vida.

AGRADECIMENTOS

Gostaria de iniciar agradecendo a Deus pela saúde, pela força e pela perseverança que me permitiram concluir esta jornada.

Aos professores da Escola de Guerra Naval, sou imensamente grata pelo conhecimento transmitido e pelas lições que levarei para a vida toda.

Um agradecimento especial a CMG (RM1-T) Chiara pelas orientações e direcionamentos que cuidadosamente me forneceu.

Ao meu orientador, CC Roberto Pimenta, meu sincero agradecimento pelas orientações, pela paciência e pelo tempo dispendido para ler e encaminhar uma direção para a melhoria deste trabalho.

À minha família, agradeço pela compreensão e pelo apoio incondicional, especialmente nos momentos em que estive ausente.

Aos colegas de turma, que se tornaram verdadeiros amigos, e a todos os meus amigos que me deram uma palavra de incentivo ao longo do caminho, sou profundamente grata. Este trabalho é fruto do esforço conjunto de todos que me acompanharam nesta caminhada.

C-Sup 2024

“Buscai primeiro o Reino de Deus e todas as coisas vos serão acrescentadas.”

Bíblia Sagrada

RESUMO

Os ataques cibernéticos têm se expandido nos últimos anos, causando impactos devastadores. Em virtude disso, a preocupação dos Estados com a defesa cibernética aumentou, levando à adoção de medidas de proteção que visam mitigar as vulnerabilidades existentes. Essas medidas têm incentivado a pesquisa e a implementação de tecnologias voltadas à proteção das informações e das infraestruturas críticas. Os Sistemas de Detecção de Intrusão (IDS) são ferramentas essenciais para monitorar e prevenir ataques cibernéticos, com o objetivo de tornar o espaço cibernético mais seguro. Soluções de inteligência artificial, especificamente aquelas baseadas em aprendizado de máquina, têm sido cada vez mais utilizadas para automatizar a detecção de ataques e enfrentar os complexos desafios da segurança cibernética. Este trabalho tem como objetivo analisar a contribuição do uso de aprendizado de máquina na implementação de sistemas IDS. Para atingir esse objetivo, foram realizadas pesquisas sobre o estado da arte que empregam técnicas de aprendizado de máquina e aprendizado profundo na implementação de sistemas de detecção de intrusão. Com base nas pesquisas analisadas, concluiu-se que o aprendizado de máquina e o aprendizado profundo oferecem caminhos promissores para a detecção de intrusões, sendo que técnicas supervisionadas, como SVM e ANN, já demonstraram sucesso significativo, enquanto as abordagens de aprendizado profundo, como CNN e RNN, oferecem ainda mais precisão e capacidade de detecção de padrões complexos.

Palavras-chave: Sistemas de Detecção de Intrusão. Aprendizado de Máquina. Defesa Cibernética.

ABSTRACT

CYBER DEFENSE: THE USE OF INTRUSION DETECTION SYSTEMS SUPPORTED BY MACHINE LEARNING TO OPTIMIZE ATTACK DETECTION.

Cyberattacks have expanded in recent years, causing devastating impacts. As a result, state concerns over cybersecurity have increased, leading to the adoption of protective measures aimed at mitigating existing vulnerabilities. These measures have encouraged research and the implementation of technologies focused on protecting information and critical infrastructures. Intrusion Detection Systems (IDS) are essential tools for monitoring and preventing cyberattacks, with the goal of making cyberspace safer. Artificial intelligence solutions, specifically those based on machine learning, have been increasingly used to automate attack detection and address the complex challenges of cybersecurity.This work aims to analyze the contribution of machine learning in the implementation of IDS systems. To achieve this goal, state-of-the-art research employing machine learning and deep learning techniques in the implementation of intrusion detection systems was conducted. Based on the analyzed research, it was concluded that machine learning and deep learning offer promising paths for intrusion detection, with supervised techniques such as SVM and ANN already demonstrating significant success, while deep learning approaches, such as CNN and RNN, offer even greater accuracy and the ability to detect complex patterns.

Keywords: Intrusion Detection System. Machine learning. Cyber Defense.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AM	Aprendizado de Máquina
ANN	Artificial Neural Network
AIDS	Anomaly-based Intrusion Detection Systems
CNN	Convolutional Neural Network
ECiber	Espaço Cibernético
END	Estratégia Nacional de Defesa
EUA	Estados Unidos da América
EGN	Escola de Guerra Naval
DARPA	Defense Advanced Research Projects Agency
DCiber	Defesa Cibernética
DoS	Denial of Service
DDoS	Distributed Denial of Service
DMDC	Doutrina Militar de Defesa Cibernética
DNN	Deep Neural Network
CNTM	Controle Naval do Tráfego Marítimo
IA	Inteligência Artificial
ICS	Industrial Control System
IDS	System detection System
HIDS	Host Intrusion Detection System
NIDS	Network Intrusion Detection System
PNSI	Política Nacional de Segurança da Informação
RNN	Recurrent Neural Network
RF	Random Forest
SI	Segurança da Informação
SIC	Segurança da Informação e Comunicação
SIDS	Signature-based Intrusion Detection System
TIC	Tecnologia da Informação e Comunicação

SUMÁRIO

1 INTRODUÇÃO

A Interligação de sistemas e infraestruturas de redes através da internet simplificou as comunicações, diminuiu as distâncias e proporcionou a realização de diversas facilidades em variados setores. O ambiente onde ocorre essa interconexão global, chamado de espaço cibernético[1] (ECiber), muito se expandiu nos últimos anos. Proporcionalmente à expansão do ECiber, a vulnerabilidade a ataques cibernéticos[2] também expandiu, possibilitando ataques cibernéticos de impacto devastador.

Os ataques cibernéticos podem ser usados para aplicar crimes de roubo de informações e extorsão, como também tem o poder de afetar setores críticos e paralisar operações estatais. Em maio de 2021, o ataque cibernético à Colonial Pipeline, que forçou a empresa a interromper suas operações por vários dias, foi um evento significativo que destacou a vulnerabilidade das infraestruturas críticas a ameaças digitais. Apesar de não causar letalidade diretamente ao inimigo, esses ataques podem enfraquecer de modo alarmante o inimigo ao serem direcionados a setores críticos como energia, transporte e finanças.

Indubitavelmente ataques cibernéticos podem ser utilizados como ferramentas de guerra para atingir objetivos estratégicos, causando danos substanciais sem recorrer a conflitos armados diretos. O caso Stuxnet (Zetter, 2014), é um exemplo histórico do uso do espaço cibernético para atacar um País. O vírus, desenvolvido em 2010, interrompeu o enriquecimento de Urânio das centrífugas que ficavam concentradas nas instalações nucleares de Natanz, cidade do Irã, localizada na província de Isfahan. A provável intenção do ataque foi retardar o programa nuclear do Irã.

A crescente dependência de sistemas digitais suscitou a necessidade de implementar políticas robustas de Segurança da Informação, tanto para empresas privadas quanto para os Estados-Nacionais. A cibersegurança, que se refere a todo esforço realizado no intuito de fornecer proteção de dados, programas, servidores e infraestrutura de redes contra acessos ou alterações não autorizados, tornou-se uma

1espaço virtual, composto por dispositivos computacionais conectados em redes ou não, onde as informações digitais transitam, são processadas e/ou armazenadas. (Brasil, 2023)

2 Vide glossário, pág. 45.

prioridade essencial para garantir a estabilidade e a resiliência de governos e instituições públicas diante de ataques cibernéticos cada vez mais sofisticados.

A busca por medidas de proteção eficazes torna-se imperiosa à medida que os avanços tecnológicos proporcionam o desenvolvimento de estratégias mais sofisticadas de ataque. Torna-se, portanto, imperativo que a defesa acompanhe essa evolução e busque o contínuo aperfeiçoamento das estratégias de prevenção e defesa. Os sistemas de detecção de intrusão, do inglês "*Intrusion Detection System (IDS)*" surgiram para monitorar e prevenir ataques, visando tornar os sistemas conectados no espaço cibernético mais seguros, uma vez que detectam os estágios iniciais do ataque, o que permite a oportunidade de mitigação. No entanto, esse objetivo é desafiador, já que técnicas de ataque muitas vezes conseguem burlar as técnicas de proteção existentes.

Diante da crescente complexidade do panorama das ameaças cibernéticas[3], surge a demanda por tecnologias que possam detectar, investigar e tomar decisões rápidas frente a ameaças e ataques emergentes. Nesse contexto, a inteligência artificial (IA), especificamente o aprendizado de máquina (AM), tem desempenhado um papel de grande relevância.

As soluções baseadas em IA emergem como uma resposta eficaz para automatizar a detecção de ataques e lidar com os desafios complexos da segurança cibernética. Pesquisas recentes têm se concentrado em técnicas de AM, visando aprimorar a detecção de intrusões e a prevenção de ataques. Essas técnicas têm obtido bons resultados na identificação e mitigação de ameaças em tempo real, aprimorando significativamente a segurança dos sistemas conectados.

Diante do exposto, na presente monografia objetiva-se analisar a contribuição das técnicas de AM na implementação dos sistemas de detecção de intrusão e seus principais desafios. Para isso, essa pesquisa será norteada com a seguinte questão: Como a IA tem sido usada para a defesa do espaço cibernético nos últimos 5 anos, principalmente para as técnicas voltadas para a área de detecção de intrusão?

Para realização da pesquisa bibliográfica foram procurados livros, artigos e trabalhos científicos disponibilizados na Rede de Bibliotecas Integradas da Marinha (RedeBIM), na Biblioteca da Escola de Guerra Naval (EGN), no IEEE *Xplorer,* e no

3 causa potencial de um incidente indesejado que pode resultar em dano ao espaço cibernético de Interesse. (Brasil, 2023)

Google Schoolar. As palavras-chave utilizadas para Pesquisa foram: (“Aprendizado de máquina” e “Segurança cibernética”); (“Machine Learning” and “cyber security”); (“Aprendizado de máquina” e “Detecção de intrusão”) ; (“Machine Learning” and “Intrusion Detection”). Para o embasamento teórico da Inteligência Artificial, principal mente do Aprendizado de máquina em seus múltiplos aspectos, será utilizada como referência principal será Faceli et. al (2021), o qual aborda os fundamentos, principais métodos e aplicações, fornecendo uma base teórica sólida nesse campo do conhecimento. Para a compreensão dos métodos de detecção de intrusão, será usado como fonte principal Halboni et. al (2022), para a investigação e análise dos avanços recentes em Sistemas de Detecção de Intrusão com foco na utilização de Aprendizado de máquina, será utilizado como referência, entre outras fontes, o trabalho de Vanin (2022) e Halboni et. al (2022).

A fim de atender ao seu propósito, esta monografia está organizada em quatro outros capítulos. No segundo capítulo, será apresentado o esforço que tem sido realizado para defesa cibernética no Brasil e na MB, e também os conceitos técnicos de detecção de intrusão e aprendizado de máquina. No terceiro capítulo, serão vistos casos de ataques cibernéticos, em particular, os com objetivos militares. No quarto capítulo, a atual aplicabilidade do aprendizado de máquina na detecção de intrusão será abordada, com o foco na aplicação de técnicas e dos resultados já alcançados nos últimos 5 anos. Além disso, nesse capítulo também serão vistos os principais desafios e limitações enfrentados para a aplicabilidade das técnicas de AM na detecção de intrusão. Por fim, no quinto capítulo, será discorrida a conclusão.

2 REFERENCIAL TEÓRICO

Neste capítulo, serão apresentados os conceitos de defesa cibernética, detecção de intrusão e aprendizado de máquina. O objetivo é fornecer uma base sólida de conhecimento para os leitores, destacando os princípios teóricos fundamentais que facilitarão a compreensão dos temas abordados ao longo deste trabalho. A ênfase será no nivelamento conceitual, garantindo que todos os leitores possuam o entendimento necessário para acompanhar as discussões subsequentes.

2.1 Defesa cibernética no Brasil e na MB

Nosso País, nos últimos anos, tem se esforçado para fortalecer sua Defesa cibernética(D-Ciber), reconhecendo a relevância crescente da segurança digital em um mundo cada vez mais conectado. A Estratégia Nacional de Defesa (END)(Brasil, 2020a), visando ao atendimento dos Objetivos Nacionais de Defesa(OND) determinados no Plano Nacional de Defesa(PND), definiu o setor cibernético como um dos setores tecnológicos essenciais para a defesa nacional, em conjunto com os setores nuclear e espacial.

Conforme Brasil (2020a), para o Setor Cibernético, foram definidas prioridades para as tecnologias de comunicação entre as unidades das Forças Armadas a fim de garantir interoperabilidade e segurança nas operações integradas. Isso exige melhorias na Segurança da Informação, das Comunicações e na Segurança Cibernética em todas as esferas do Estado, com foco na proteção das Infraestruturas Críticas.

Com vistas a atender aos objetivos de proteção do Eciber, diversas normativas foram estabelecidas. No âmbito da União Federal, foi instituída a Política Nacional de Segurança da Informação (PNSI), com objetivos de assegurar a disponibilidade, a integridade, a confidencialidade e a autenticidade da informação a nível nacional. (Brasil,2021a)

Na PNSI, em seu Art.2, está disposto que a Segurança da Informação (SI) é composta pela Segurança Cibernética (SegCiber), pela Defesa Cibernética (Dciber), bem como pela segurança física e a proteção de dados organizacionais. A Política Cibernética de Defesa, normatizada no MD31-P-02, direciona as atividades de DCiber no Ministério da Defesa, abrangendo a Guerra Cibernética (GCiber) nos níveis estratégico, operacional e tático, aplicando-se a todos os componentes militares e às entidades envolvidas nessas atividades. (Brasil,2021a)

A Doutrina Militar de Defesa Cibernética (DMDC) define os princípios da Defesa Cibernética contribuindo para a atuação conjunta das Forças Armadas na defesa do Brasil no ECiber, e formaliza a responsabilidade de cada força na proteção e defesa cibernética dos seus ativos de informação.(Brasil,2021a)

A Marinha do Brasil, no seu âmbito, ao assumir a sua responsabilidade de proteção dos seu ativos de informação, atua tanto na Defesa Cibernética quanto na Guerra Cibernética.

Brasil(2019) estabeleceu em seu objetivo 8 “Desenvolver a capacidade cibernética”, que diz respeito à preparação do Poder Naval para os conflitos no ECiber. Sendo necessário, para isso, a integração das forças navais com os setores de guerra cibernética da MB e demais Forças. Neste documento é definido que as ações de defesa cibernética devem atuar em três vertentes, a saber: proteção dos ativos de Tecnologia da Informação e Comunicações; a exploração cibernética e no ataque cibernético(Brasil, 2019).

Nesse sentido, é enfatizado por Brasil (2019) a importância a proteção das infraestruturas críticas de TIC da MB com foco especial nas forças navais em operação, o que inclui a implementação de medidas de segurança para evitar ataques cibernéticos que possam comprometer a eficiência ou a segurança das operações navais. Ademais, a exploração cibernética visa a melhoria da tomada de decisões, proporcionando uma vantagem estratégica ao conhecer melhor as capacidades e intenções do inimigo. E por fim, o ataque cibernético, que visa a utilização de meios cibernéticos para degradar, corromper ou neutralizar os sistemas e ativos de TIC, é também de grande importância ao prover apoio às operações de guerra naval, buscando enfraquecer ou paralisar as capacidades operacionais do inimigo.

Além disso, o fortalecimento da Segurança da Informação e Comunicações (SIC), definido como objetivo setorial associado ao desenvolvimento da capacidade cibernética da MB, norteia inúmeras ações estratégicas já implementadas e em desenvolvimento sob a responsabilidade da Diretoria de Comunicações e Tecnologia da Informação da Marinha (DCTIM). Uma das ações para o fortalecimento do SIC é justamente a prevenção de ataques oriundos das redes às quais os sistemas estão conectados. (Brasil, 2024).

O IDS, que será visto na próxima seção, contribui para a prevenção de ataques ao monitorar o tráfego de rede e identificar atividades suspeitas ou anômalas em tempo real. Ao detectar padrões que podem indicar tentativas de invasão ou comportamentos maliciosos, esses sistemas alertam os administradores de segurança, permitindo que ações preventivas sejam tomadas antes que um ataque cause

danos significativos. Dessa forma, o IDS atua como uma camada de defesa essencial, complementando outras medidas de segurança e reduzindo o tempo de resposta a ameaças.

2.2 Os sistemas de Detecção de Intrusão

Para mitigar os riscos associados às ameças computacionais, principalmente aos sistemas conectados ao ECiber, várias medidas de segurança precisam ser implementadas. Dentre elas estão: monitoramento contínuo, atualização dos sistemas, políticas de segurança, educação e treinamento aos usuários do sistema. Os sistemas de detecção de intrusão enquadram-se na medida de monitoramento, com o objetivo de monitorar e prevenir ameaças.

Intrusão em um sistema computacional pode ser definida como um desvio do comportamento normal desse sistema, por meio de acesso não autorizado (Gupta,2020). A intrusão visa, fundamentalmente, o roubo ou a alteração de informações, com o propósito de causar danos a indivíduos ou organizações conectadas a esses sistemas. Tais ações podem comprometer a integridade, a confidencialidade e a disponibilidade dos dados e serviços, resultando em diversos tipos de impactos adversos.

Entre as consequências da intrusão em sistemas computacionais destacam-se: Negação do serviço, do inglês "*Denial of Service (DoS)*[4]" e Negação de serviço Distribuída, do inglês *"Distributed Denial of Service(DDoS)*[5]". Os sistemas de detecção de intrusão, do inglês "*Intrusion Detection System (*IDS)" são ferramentas que monitoram o comportamento e são utilizadas para detectar anormalidades, bem como prevenir essas atividades maliciosas.

2.2.1 Classificação de IDS

No que diz respeito à técnica de implementação, os Sistemas de Detecção de intrusão(IDS) podem ser baseados em anomalia ou em assinatura. Os sistemas IDS

4 Vide Glossário, pág. 46
5 Vide Glossário, pág. 46

por anomalia, do inglês ”*Anomaly-based Intrusion Detection Systems (AIDS)*", têm como objetivo identificar atividades que desviam do padrão, ou seja, conseguem detectar comportamentos anômalos por meio da comparação com uma base de conhecimento.

Para construir um sistema IDS baseado em anomalia o primeiro passo é a coleta de dados. Esses registros incluem logs de servidores, *firewalls*[6], dados de aplicação, entre outros; informações de atividades dentro de aplicativos; e comportamento de usuários, como atividades de login e acessos a arquivos, por exemplo.

Com base nestes dados, é criado um perfil que representa o comportamento convencional dessa rede ou sistema. Esse perfil pode ser criado a partir de uma análise estatística, que define o comportamento por meio de taxas de tráfego da rede ou padrões de acesso a arquivos, por exemplo. Além disso, o perfil pode ser desenvolvido por modelos de comportamento ou por Aprendizado de máquina.

Nos perfis criados por modelo de comportamento, os padrões são estabelecidos por meio de modelos específicos de comportamento do sistema, baseado em regras ou heurísticas. Já a utilização de aprendizado de máquina permite que os padrões sejam aprendidos com base no comportamento normal, através da coleta de dados. O pressuposto para os IDS por anomalias é que o comportamento anômalo difere do comportamento típico. Ao detectar uma atividade que se desvia significativamente do comportamento normal estabelecido no perfil base, o IDS identifica uma intrusão.

Um IDS baseado em assinatura, do inglês “*Signature-based Intrusion Detection System*)(SIDS)”, funciona de forma semelhante a um antivírus, identificando ataques e intrusões por meio do tráfego da rede ou dos arquivos do sistema contra um banco de dados de assinaturas conhecidas de ataques e vulnerabilidades. O SIDS mantém um banco de dados atualizado com essas assinaturas, que é frequentemente renovado pelos fornecedores do IDS à medida que novos tipos de ataques são descobertos. O tráfego de rede, arquivos de sistema, ou ambos, dependendo do tipo de IDS, é monitorado continuamente. O sistema inspeciona pacotes de dados, analisando-os para detectar correspondências com assinaturas conhecidas. Quando um pacote de dados ou um evento corresponde a uma assinatura armazenada em

6 Vide Glossário. Pág. 47

seu banco de dados, o IDS gera um alerta.

Os sistemas IDS podem ser instalados na rede ou no próprio host. Os sistemas instalados nas máquinas (hosts) são chamados de HIDS, do inglês, *Host Intrusion Detection System(HIDS)*. Existem também os sistemas de detecção de intrusão instalados em rede, chamados de NIDS, do inglês, *Network Intrusion Detection System (NIDS).*

Os HIDS, instalados em hosts, verificam os arquivos que são acessados, aplicações e informações de arquivos de log do sistema operacional, ou seja, informações relativas ao tráfego local de uma estação de trabalho, por exemplo. Por sua vez, os NIDS, instalados na rede, monitoram as informações que trafegam em uma rede de computadores, incluindo pacotes de serviços, portas, requisições, dentre outras. Existem também os IDS híbridos, que combinam o melhor do HIDS com o melhor dos NIDS.

2.2.2 Vantagens e desvantagens dos tipos de IDS

Um dos principais desafios enfrentados por grande parte dos IDS existentes são a taxa de falsos alarmes e a incapacidade de detectar novas intrusões. A principal limitação dos IDS por assinatura é detectar somente ataque cuja assinatura está armazenada no banco de dados, não sendo possível a detecção de ataques do dia zero (Gupta, 2020). Ataques de dia zero é uma exploração cibernética que tira vantagem de uma vulnerabilidade de *software* desconhecida pelo fornecedor, ocorrendo antes que uma solução esteja disponível. Por ser um tipo de ataque novo, ou seja, cuja assinatura não é conhecida, não se encontra no banco de dados, portanto, não é detectada por IDS baseados em assinatura.

Por outro lado, uma vez que os IDS por anomalia buscam por comportamentos anômalos, são suscetíveis a falsos positivos (Tait, 2021). Um falso positivo ocorre quando o sistema identifica erroneamente uma atividade legítima como uma ameaça ou ataque, gerando um alerta de segurança desnecessário. Outra desvantagem desse tipo de IDS é a adaptação em ambientes dinâmicos, ou seja, onde o usuário exige muitas mudanças ao longo do tempo, tornando-se mais difícil determinar o comportamento padrão (Tait, 2021).

2.3 Inteligência artificial e o Aprendizado de máquina

As soluções computacionais anteriores à IA, embora muito úteis para resolver determinados problemas, não eram adequadas para resolver problemas que seres humanos resolvem com facilidade, tais como reconhecer uma pessoa pela sua voz ou pelo seu rosto, ou combinar conhecimentos obtidos por meio das experiências passadas. A partir dos anos 1970, houve uma expansão no uso da IA. Inicialmente, os problemas eram tratados por meio de sistemas especialistas, que eram programas utilizados para aquisição de conhecimento de um dado domínio, como a medicina, por exemplo, e que buscavam conhecer as regras que eram utilizadas para a tomada de decisão por meio de entrevistas.(Faceli, 2021)

A evolução computacional, aliada ao aumento no volume de dados e à crescente complexidade dos problemas a serem solucionados computacionalmente, proporcionou o desenvolvimento de ferramentas computacionais mais sofisticadas e independentes da intervenção humana. Essas ferramentas se baseiam predominantemente no Aprendizado de Máquina (AM), uma subárea da Inteligência Artificial (IA), que é acompanhada por outras subáreas como Processamento de Linguagem Natural, Representação do Conhecimento, Raciocínio Automático, Visão Computacional e Robótica. A integração dessas áreas da IA tem permitido que tarefas computacionais desafiadoras se tornem triviais e amplamente aplicadas em diversos setores.

2.3.1 Tipos de Aprendizado de Máquina

O objetivo do Aprendizado de Máquina é fazer com que o computador aprenda uma tarefa com base na experiência acumulada. Os algoritmos de Aprendizado de Máquina (AM) têm sido amplamente aplicados em várias tarefas, que podem ser classificadas como Preditivas e Descritivas.

Em tarefas descritivas, algoritmos de AM identificam padrões nos atributos preditivos de um conjunto de dados sem a orientação de um supervisor, sendo, portanto, de aprendizado não supervisionado. Uma de suas principais funções é agrupar dados, identificando grupos de objetos semelhantes entre si.

Nas tarefas preditivas, algoritmos são aplicados a um conjunto de dados rotulados para prever o valor de um atributo alvo com base nos valores dos atributos preditivos de um novo objeto. Esse tipo de aprendizado é chamado supervisionado Uma das principais características do aprendizado supervisionado é usar experiências anteriores para produzir resultados. A aprendizagem supervisionada pode ser dividida em dois tipos de modelos: classificação e regressão.

Os modelos de classificação são utilizados para categorizar dados em grupos específicos. A partir de um conjunto de dados, o modelo identifica a categoria à qual cada entrada pertence, com base em suas características. O modelo é treinado com dados de entrada e saída rotulados, para aprender as características dos dados de entrada e, assim, classificá-los corretamente. Por outro lado, os modelos de regressão são utilizados para prever resultados contínuos. O modelo é treinado para entender a relação entre variáveis independentes e uma variável dependente. A regressão é empregada para identificar padrões e relações em conjuntos de dados, que podem ser aplicados a novos dados para fazer previsões.

Nas tarefas descritivas, os algoritmos de Aprendizado de Máquina extraem padrões dos valores dos atributos preditivos de um conjunto de dados, sem a presença de um supervisor externo. Esses algoritmos são classificados como de aprendizado não supervisionado. Uma das principais funções deste grupo é realizar o agrupamento de dados, identificando grupos de objetos que são similares entre si dentro do conjunto de dados. Assim, no aprendizado não supervisionado, o modelo melhora-se por si só, ao descobrir padrões e informações do conjunto de dados utilizado.

Por fim, o aprendizado semi-supervisionado representa a junção dos dois tipos de aprendizado de máquina: supervisionado e não supervisionado. As técnicas e métodos utilizados nos aprendizados supervisionado e não supervisionado podem ser aplicados ao conjunto de dados.

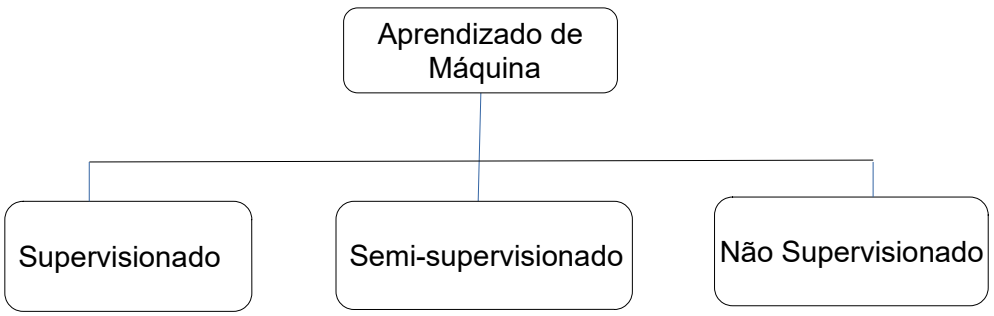

Fig1. Hierarquia de aprendizado de máquina

2.3.2 Técnicas de AM utilizadas para detecção de intrusão

A aplicação do aprendizado de máquina pode ser eficaz tanto na identificação de ataques já conhecidos quanto na detecção de ameaças desconhecidas, desde que o modelo tenha sido devidamente treinado. Os modelos de classificação são especialmente valiosos nessa tarefa. Por exemplo, ao analisar o tráfego que ingressa em nossa rede, um modelo de classificação bem treinado pode diferenciar entre tráfego normal e tráfego anômalo. Algoritmos de classificação tais como árvores de decisão, Floresta Aleatória, Máquina de vetor de suporte e rede Neural são muito comuns para este tipo de tarefa (Vanin, 2022).
A seguir são apresentados o funcionamento de alguns desses algoritmos.

- FLORESTA ALEATÓRIA

O método Floresta Aleatória (Random Forest – RF), é uma técnica de aprendizado de máquina que utiliza múltiplas árvores de decisão para classificação e regressão. Faz parte dos algoritmos de aprendizado por comitê, ou "*ensemble learning*[7]". (Breiman, 2001)

- MÁQUINA DE VETOR DE SUPORTE

A Máquina de vetor de suporte, do inglês *"Support Vector Machine (SVM)",* é uma técnica de aprendizado usada para classificação e regressão. Seu objetivo é encontrar um hiperplano que separe eficientemente as diferentes classes de dados no espaço de características. (Gerón, 2021)

- REDE NEURAL ARTIFICIAL

As redes neurais artificiais (RNA), do inglês "*Artificial Neural Network*" são modelos computacionais inspirados na estrutura e funcionamento do cérebro humano. Elas consistem em unidades de processamento interconectadas chamadas neurônios artificiais. Os neurônios artificiais da RNA são inspirados nos neurônios biológicos; cada neurônio recebe sinais, processa esses sinais e transmite a

7 Vide glossário, pág. 46.

informação para outros neurônios. Uma rede neural funciona com várias camadas de neurônios. A primeira camada recebe os dados, as camadas intermediárias processam essas informações, e a última camada entrega o resultado. Cada neurônio está conectado a outros neurônios nas camadas adjacentes, e essas conexões têm pesos que influenciam o processo de aprendizado. (Gerón, 2021)

2.3.3 Avaliação do desempenho dos IDS que utilizam AM

Para avaliar os IDS que utilizam aprendizado de máquina são utilizadas métricas de desempenho. Essas métricas servem para analisar os resultados da matriz de confusão, que é uma tabela utilizada para exibir o desempenho de um modelo de classificação. A matriz de confusão compara os valores reais dos dados com os valores previstos pelo modelo de aprendizado de máquina.

Para descrever uma matriz de confusão, utilizamos os seguintes termos:

- Verdadeiro Positivo (VP), do inglês *"True Positive (TP)"*: Quando uma amostra de ataque é corretamente identificada como um ataque pelo modelo.
- Verdadeiro Negativo (VN), do inglês *"True Negative (TN)"*: Quando uma amostra normal (tráfego legítimo) é corretamente identificada como tal pelo modelo.
- Falso Positivo (FP), do inglês *"False Positive (FP)"*: Quando uma amostra normal é incorretamente identificada como um ataque pelo modelo.
- Falso Negativo (FN), do inglês "*False Negative (FN)*". Quando uma amostra de ataque é incorretamente identificada como tráfego normal pelo modelo.

Tabela1 – Métricas para medir o desempenho do IDS – Baseado em Vanin (2022).

		Classe Predita (Modelo)	
		Normal	**Ataque**
Classe Atual (Dados)	**Normal**	VN	FP
	Ataque	FN	VP

As principais métricas de avaliação utilizadas para classificar os modelos usados para IDS são:

- **Precisão** - A taxa de precisão corresponde a proporção de verdadeiros positivos (VP) entre todas as predições positivas, ou seja, corresponde a razão entre as amostras de ataques corretamente previstas e todas as amostras de ataques previstas.
- **Acurácia** - Corresponde à razão entre as classes corretamente identificadas e todas as amostras.
- **Recall** - Corresponde à razão entre as amostras de ataques corretamente previstas e todas as amostras que correspondem a um ataque. Esta métrica também é conhecida como Taxa de Detecção.
- **Medida F1** - Corresponde à média harmônica da Precisão e do Recall. Fornece uma melhor avaliação do sistema ao mostrar a diferença entre as duas métricas, verificando se a solução está balanceada. É também conhecida como F-Score ou F1-Score.

2.4 Conclusão parcial

Neste capítulo foram vistos alguns conceitos sobre os Sistemas de Detecção de intrusão, aprendizado de máquina e métricas de avaliação com a finalidade de melhor entendimento dos capítulos posteriores. Foi discutida a diferença entre os IDS baseados em assinatura e em anomalia, evidenciando as limitações de cada abordagem, como a incapacidade dos IDS por assinatura de detectar ataques de dia zero e a susceptibilidade dos IDS por anomalia a falsos positivos.

Além disso, foi explorado o funcionamento das técnicas Florestas Aleatórias, Máquinas de Vetores de Suporte e Redes Neurais Artificiais, muito utilizadas em de aprendizado de máquina. Por fim, foram vistos os termos utilizados para a formação das métricas de avaliação, como precisão, acurácia, recall e Medida F1, as quais são utilizadas para medir a eficácia dos modelos de aprendizado de máquina na prática.

3 AÇÕES CIBERNÉTICAS EM GUERRA

Neste capítulo, serão analisados casos recentes de ataques cibernéticos, com ênfase especial nos que têm objetivos militares. Na seção 3.1, serão discutidas as ações cibernéticas que antecederam e prepararam o terreno para as ofensivas de guerra. Já na seção 3.2, serão exploradas as ações cibernéticas com finalidades militares, sejam elas exclusivamente virtuais ou combinadas com operações militares tradicionais. O objetivo principal deste capítulo é apresentar episódios significativos de ações cibernéticas com propósitos militares, destacando a importância de fortalecer as estratégias de defesa cibernética diante dessas novas ameaças.

3.1 Ações cibernéticas na Pré - Guerra

Clarke e Knake(2015) relatam um episódio significativo sobre as ações cibernéticas realizadas pelos Estados Unidos antes da guerra com o Iraque em 2003. Conforme descrito pelos autores, durante a segunda guerra contra o Iraque, os Estados Unidos implementaram uma estratégia cibernética significativa antes do ataque inicial dos caças-bombardeiros. A rede militar privada e segura do Iraque, antes mesmo do início das hostilidades, já havia sido comprometida, fato que os americanos comunicaram aos militares iraquianos. Conforme descrito por Clarke e Knake(2015), pouco antes da guerra, milhares de oficiais iraquianos receberam e-mails provenientes do sistema do Ministério da Defesa do Iraque. Embora o conteúdo completo desses e-mails nunca tenha sido divulgado, fontes confiáveis revelaram que a mensagem essencialmente instruía os militares iraquianos a não resistirem ao ataque iminente dos EUA. A mensagem, supostamente enviada pelo Comando Central dos Estados Unidos, indicava que os americanos não desejavam prejudicar os militares iraquianos, mas sim remover Saddam Hussein e seus filhos do poder. Os oficiais eram incentivados a abandonar seus tanques e veículos blindados, posicionando-os de maneira organizada e retornando para casa. Muitos oficiais iraquianos seguiram essas instruções. Como resultado, as tropas americanas encontraram tanques de várias unidades estacionados em filas fora das bases, facilitando a destruição organizada por aviões dos EUA. Alguns comandantes

iraquianos até dispensaram suas tropas antes do início da guerra, abandonando suas posições e tentando retornar para suas casas.

Apesar da disposição do governo Bush em invadir a rede iraquiana para uma campanha psicológica, eles optaram por não atacar os ativos financeiros de Saddam Hussein. Mesmo tendo capacidade para tal, os advogados do governo temiam que a invasão de contas bancárias fosse vista como uma violação do direito internacional, criando um precedente perigoso. Além disso, havia preocupações de que os ataques cibernéticos aos bancos pudessem ter efeitos colaterais imprevistos, como afetar contas erradas ou derrubar instituições inteiras, incluindo bancos dos EUA.

Outro ataque cibernético significativo, também relatado na obra de Clarke e Knake(2015), trata-se do ocorrido contra uma instalação nuclear na Síria. Em 6 de setembro de 2007, um ataque cibernético e aéreo coordenado foi realizado por Israel contra uma instalação na Síria. Localizada a 120,7 quilômetros da fronteira turca, a construção era atribuída a um projeto secreto envolvendo a Coreia do Norte. À meia-noite, a tranquilidade foi interrompida por uma série de explosões que destruíram completamente o complexo, com aeronaves F-15 Eagle e F-16 Falcon retornando à Israel sem sofrer danos. A reação inicial foi de silêncio tanto por parte de Israel quanto da Síria, que alegou que o prédio destruído era apenas uma estrutura vazia. Apenas a Coreia do Norte protestou abertamente.

A imprensa ocidental começou a divulgar que a instalação era uma fábrica de armas nucleares. No entanto, surgiram dúvidas sobre a veracidade dessas alegações, com algumas fontes sugerindo que o prédio poderia estar relacionado ao programa de mísseis da Síria ou que Israel estava enviando um aviso ao Irã. Independentemente das teorias, o ataque destacou a capacidade de guerra cibernética de Israel, que conseguiu desativar a defesa aérea síria, deixando o espaço aéreo vulnerável ao ataque.

Os relatórios subsequentes revelaram que a CIA produziu um vídeo confirmando que a instalação era, de fato, um reator nuclear em construção, apoiando a versão israelense. A Agência Internacional de Energia Atômica (IAEA) também encontrou vestígios de materiais radioativos no local, reforçando a acusação de que a Síria estava desenvolvendo armas nucleares com ajuda norte-coreana. No entanto, o mistério sobre como Israel conseguiu neutralizar a defesa

aérea síria levantou questões importantes sobre as capacidades cibernéticas empregadas no ataque.

Conforme Clarke e Knake(2015), três teorias principais explicam como Israel pode ter conseguido esse feito: o uso de um drone furtivo para interferir nos sistemas de radar, a inserção de um backdoor no código dos computadores de defesa aérea sírios, ou a infiltração física de agentes que manipulassem a infraestrutura de fibra ótica da rede de defesa.

3.2 Ataques totalmente cibernéticos e ataques combinados

Conforme relatado por Clarke e Knake(2015), em abril de 2007, a Estônia enfrentou um dos maiores ataques cibernéticos já registrados até então. A origem do conflito remonta à decisão do governo estoniano de remover uma estátua de bronze do soldado do Exército Vermelho, símbolo da antiga ocupação soviética, que gerou forte reação da comunidade russa local e de Moscou. Em resposta à realocação da estátua, diversos ataques de negação de serviço distribuídos (DDoS) foram lançados contra o país, sobrecarregando e paralisando servidores essenciais.
Esses ataques DDoS utilizaram botnets[8], redes de computadores zumbis infectados com malware[9], tipo bot[10] que inundaram os servidores estonianos com requisições, tornando-os inacessíveis. Os alvos incluíam sites de bancos, jornais, e serviços governamentais, afetando seriamente a infraestrutura digital da Estônia. A gravidade dos ataques chamou a atenção internacional e levou a OTAN a enviar especialistas para auxiliar na defesa cibernética.

Os ataques não foram resolvidos rapidamente, persistindo por semanas e demonstrando a capacidade de adaptação das botnets. Investigações rastrearam a origem dos comandos das botnets até a Rússia, gerando tensões diplomáticas, já que o governo russo negou envolvimento direto. Apesar disso, a recusa de Moscou em cooperar com a investigação estoniana indicava uma complexa dimensão política no ciberespaço.

8 Vide Glossário, pág. 45.
9 Vide Glossário, pág. 47.
10 Vide Glossário, pág 45.

Em 2008, foi criado pela OTAN, um centro de defesa cibernética em Tallinn, Estônia, em resposta ao ataque cibernético de 2007, mas a utilidade desse centro foi limitada durante o conflito entre a Geórgia e a Rússia, em 2008. Esse conflito histórico entre a Rússia e a Geórgia resultou na independência georgiana em 1991, mas em 2008, a situação escalou quando rebeldes na Ossétia do Sul provocaram a Geórgia com ataques de mísseis. A resposta georgiana levou a uma invasão, à qual a Rússia respondeu rapidamente com uma ofensiva militar e cibernética coordenada, incluindo ataques DDoS que paralisaram a infraestrutura da Geórgia.

Os ataques cibernéticos russos visaram incapacitar as comunicações georgianas, bloqueando acesso a meios de comunicação internacionais e derrubando sites governamentais. A Geórgia teve que migrar seus *sites* para servidores fora do país para manter alguma funcionalidade. As ações cibernéticas russas foram caracterizadas por uma alta coordenação e sofisticação, indicando provável envolvimento governamental, apesar das alegações russas de que eram ataques espontâneos de patriotas.

Os eventos na Geórgia revelaram a capacidade russa de combinar ataques cibernéticos com operações militares convencionais, demonstrando um novo tipo de guerra híbrida. A resposta internacional foi limitada, com um acordo de paz mediado pelo presidente francês Nicolas Sarkozy, mas a Rússia manteve seu controle sobre as regiões separatistas de Ossétia do Sul e Abkházia. A atitude russa frente ao ciberespaço mostrou um desprezo pelas normas internacionais de paz cibernética.

O ataque cibernético à Geórgia mostrou a vulnerabilidade das infraestruturas nacionais a operações cibernéticas bem coordenadas. A Geórgia tentou várias táticas defensivas, como bloquear tráfego da Rússia, mas os atacantes se adaptaram rapidamente, redirecionando os ataques para parecerem originados de outros países. O ataque também teve impacto econômico, isolando o setor bancário georgiano e causando paralisações significativas.

Conforme disposto em Clarke e Knake (2015), foram feitas comparações com o ataque anterior à Estônia, sugerindo que a Rússia estava testando suas capacidades cibernéticas em diferentes cenários. No entanto, ao contrário do ataque à Estônia, o ataque à Geórgia foi um prelúdio para uma invasão militar direta, destacando a possibilidade de ataques cibernéticos prepararem o terreno para

operações militares. Diante disto, a comunidade internacional reconheceu a necessidade de fortalecer suas defesas cibernéticas diante dessa nova forma de ameaça.

3.3 Conclusão parcial

Os episódios descritos por Clarke e Knake sobre as ações realizadas pelos EUA e sobre o ataque de Israel à instalação nuclear da Síria demonstram como as operações cibernéticas se tornaram uma parte de elevada importância na execução das estratégias militares modernas, contribuindo significativamente para as ofensivas militares. No caso do ataque cibernético dos EUA contra o Iraque em 2003, a habilidade de comprometer a rede militar privada e segura do país antes mesmo do início das hostilidades representou uma vantagem tática significativa. A campanha psicológica que seguiu, na qual os oficiais iraquianos foram instruídos a não resistirem e a abandonarem seus postos, facilitou a invasão e a destruição organizada dos equipamentos militares iraquianos. Estas operações, que incluíam a infiltração de redes de comunicação e sistemas de controle iraquianos, disseminação de desinformação para confundir e desorganizar as forças armadas não só reduziu a resistência durante a invasão, mas também exemplificou o poder de manipulação e desmoralização que as operações cibernéticas podem exercer sobre um exército adversário.

De maneira semelhante, o ataque coordenado de Israel contra a instalação nuclear na Síria em 2007 ressaltou a importância das capacidades cibernéticas para neutralizar defesas inimigas. A habilidade de Israel em desativar a defesa aérea síria, permitindo um ataque aéreo sem oposição, demonstrou como a guerra cibernética pode complementar operações tradicionais, ampliando significativamente o alcance e a eficácia das ações militares. Indubitavelmente, esse tipo de operação exemplifica como ataques cibernéticos podem ser utilizados para complementar as ofensivas, criando um aumento significativo para a eficácia da missão. Ambos os casos sublinham a crescente interdependência entre operações cibernéticas e militares, destacando que a superioridade cibernética pode ser tão decisiva quanto a superioridade aérea ou terrestre em conflitos modernos.

Os ataques cibernéticos à Estônia em 2007 e à Geórgia em 2008 demonstram a evolução e a eficácia dos ciberataques como ferramentas de ofensivas militares. No caso da Estônia, os ataques de negação de serviço distribuídos (DDoS) paralisaram setores essenciais do país, destacando a vulnerabilidade das infraestruturas digitais nacionais a ataques cibernéticos organizados, principalmente em nações altamente conectadas. A resposta internacional, especialmente da OTAN, revelou a importância da cooperação em defesa cibernética, resultando na criação do Centro de Defesa Cibernética em Tallinn. Essa medida foi um marco importante, mas sua eficácia em conflitos subsequentes mostrou-se limitada, evidenciando a necessidade contínua de aprimoramento das estratégias de defesa cibernética.

Já no conflito entre Rússia e Geórgia, a combinação de ataques cibernéticos com operações militares convencionais evidenciou uma nova dimensão de guerra híbrida. Os ataques DDoS coordenados contra a Geórgia visaram desestabilizar a infraestrutura de comunicação e causar caos antes e durante a invasão militar. Este episódio destacou a capacidade de ciberataques prepararem o terreno para operações militares, criando uma vantagem estratégica significativa. A resposta internacional limitada e a subsequente manutenção do controle russo sobre regiões separatistas ilustraram a complexidade das relações internacionais frente a essa nova forma de ameaça. Em suma, os ataques à Estônia e à Geórgia demonstram como os ciberataques podem ser eficazes em desestabilizar nações e apoiar ofensivas militares. A resposta da comunidade internacional, ainda que significativa, mostrou a necessidade de contínua adaptação e fortalecimento das defesas cibernéticas.

Esses eventos sublinharam a urgência de estabelecer medidas de segurança cibernética robustas para proteger contra ameaças semelhantes no futuro, além de evidenciar a necessidade de estabelecer normas internacionais claras para a paz cibernética e a cooperação global em defesa contra ameaças cibernéticas, prevenindo assim o uso de ciberespaço como campo de batalha.

4 SISTEMAS IDS COM APRENDIZADO DE MÁQUINA

A crescente complexidade e o volume de dados trafegados nas redes modernas tornam os métodos tradicionais de detecção de intrusão insuficientes. As técnicas de aprendizado de máquina (AM) permitem que os IDS evoluam, aprendendo continuamente a partir de novos dados e adaptando-se a novos tipos de ataques.

Os IDS são ferramentas essenciais na defesa de redes de computadores contra acessos não autorizados e atividades maliciosas. Conforme abordado no Capítulo 2, existem diversos tipos de IDS. Com o avanço das técnicas de AM, essas ferramentas têm se tornado cada vez mais eficazes. De acordo com Halboni et al. (2022), o objetivo do uso de técnicas de aprendizado de máquina é criar IDS com maior precisão e menor exigência de intervenção humana. No entanto, veremos que a construção de IDS baseados em anomalias (AIDS), que permitem a identificação de ataques de dia zero, é uma tarefa desafiadora, uma vez que os conjuntos de dados utilizados para o treinamento dos modelos de aprendizado de máquina não são compostos por instâncias de ataques reais e recentes.

Neste capítulo, serão apresentados os trabalhos recentes que investigaram essa área, os resultados da aplicação dessas técnicas utilizando AM, a importância dos dados utilizados nas soluções de IDS, bem como os desafios da implementação. Por fim, será apresentada uma breve conclusão desses estudos, destacando o aprendizado obtido no estudo.

4.1 A importância do Banco de dados para soluções de IDS

Em aprendizado de máquina, uma característica de grande importância são os dados os quais os modelos são treinados. A qualidade, diversidade e quantidade desses dados em Sistemas de Detecção de intrusão vão influenciar diretamente a eficácia desses sistemas, uma vez que um conjunto de dados bem representativo permite que o modelo aprenda a distinguir entre comportamentos normais e maliciosos.

Conforme Khraisat et al (2019), o primeiro esforço para criar um conjunto de dados IDS foi feito pela DARPA(Defense Advanced Research Projects Agency) em 1998, resultando no conjunto de dados KDD98. Este conjunto foi usado como base para o KDD Cup99, desenvolvido para a Terceira Competição Internacional de Descoberta de Conhecimento e Ferramentas de Mineração de Dados em 1999. Apesar de sua importância, ambos os conjuntos foram criticados por não refletirem condições reais e estarem desatualizados quanto a ataques recentes de *malware*.

O NSL-KDD foi desenvolvido em 2009 a partir do KDD Cup99 para resolver problemas de redundância e melhorar a representatividade dos dados. O principal problema no conjunto de dados KDD é a enorme quantidade de pacotes duplicados. A grande quantidade de instâncias duplicadas no conjunto de treinamento pode tornar os métodos de aprendizado de máquina tendenciosos para instâncias normais, dificultando o aprendizado de instâncias irregulares, que são mais prejudiciais ao sistema de computador(Khraisat et al.,2019).

Segundo Vanin et.al (2022), foi observado que para algumas soluções, as quais utilizaram conjunto de dados recentes, a precisão do sistema IDS diminuiu em comparação com conjunto de dados mais antigos, quando a precisão era excelente. Foi observado também pelos autores, outro ponto de preocupação relacionado às soluções estudadas: a sua ineficiência em detectar classes de ataque específicas que possuem menos amostras em seu conjunto de dados. Isso ocorre em virtude de um desequilíbrio de classes no conjunto de dados, que resulta em uma taxa de detecção mais baixa para essas classes, o que é um grande problema pois essas classes menores podem ser ataques de dia zero.

Conforme Macas et. al (2022) e Vanin et.al (2022), COPAKDD1999 e NSL-KDD são os conjuntos mais comumente usados para avaliar métodos de detecção de intrusão. Vanin et. al (2022) destaca que a qualidade do conjunto de dados é um fator importante que determina o quão eficiente será um IDS; segundo esse trabalho 56% dos IDS propostos foram testados usando KDDCup99 e NSL-KDD, ambos conjuntos de dados bem antigos.

Outros bancos de dados também utilizados na construção de IDS, porém menos utilizados são:

- O CIC-IDS 2017 - foi criado pelo Canadian Institute for Cybersecurity, inclui tráfego de rede realista e diversos tipos de ataques cibernéticos como DoS, DDoS, infiltração, *port scanning*, e ataques de força bruta.

- Kioto - coletado em 2006 pela Kyoto University em colaboração com a NEC Corporation. Contém tráfego de rede real, incluindo informações sobre origem, destino, tamanho do pacote e tempo de conexão.

- UNSW-NB15 - Criado pela University of New South Wales, dados de 2015, consiste em tráfego de rede simulado com atividades normais e nove tipos de ataques cibernéticos. Inclui uma ampla gama de características que facilitam a construção e avaliação de modelos de IDS.

-WSN-DS - Dados de 2011. Criado para redes de sensores sem fio, inclui registros de tráfego de rede coletados de simulações de ataques como *sinkhole*, *blackhole*, *flooding*, e outras anomalias. Útil para avaliar a performance de IDS em ambientes de redes de sensores.

-UNB-ISCX 2012 – Dados de 2012. Desenvolvido pelo Information Security Centre of Excellence (ISCX) na University of New Brunswick, contém tráfego de rede simulado que imita um ambiente corporativo realista, incluindo tráfego legítimo e diversos tipos de ataques.

4.2 Aplicações recentes das técnicas de Aprendizado de máquina e aprendizado profundo em IDS

Nos últimos anos, o uso de aprendizado de máquina em sistemas de detecção de intrusão tem se mostrado promissor, com diversos estudos apresentando resultados significativos. Nesta seção, analisaremos o aprendizado extraído de trabalhos que aplicaram AM e aprendizado profundo, destacando seus respectivos resultados.

Modelos como SVM e Floresta Aleatória têm sido amplamente utilizados para a categorização binária em IDS. Em um estudo citado por Alboni et al. (2022), o SVM, aplicado para a categorização binária de intrusões, apresentou um desempenho superior durante o treinamento em comparação ao Floresta Aleatória, que, por sua vez, obteve melhores resultados na fase de testes. Esse estudo

evidenciou que o desempenho de um classificador pode variar conforme o conjunto de dados e os atributos utilizados.

Outro trabalho relevante, também citado por Alboni et al. (2022), explorou um modelo híbrido de IDS combinando Árvore de Decisão, *Naive Bayes* e Floresta Aleatória para classificar ataques no conjunto de dados NSL-KDD. O modelo apresentou alta precisão na detecção de ataques DoS usando o algoritmo Floresta Aleatória, superando outros modelos híbridos em diversas categorias de ataque, como DoS, Probe, U2R e R2L.

A pesquisa de Alboni et al. (2022) também menciona o uso de Redes Neurais Artificiais (RNAs) para detectar tráfego malicioso. Nesse caso, as RNAs foram treinadas com dados variados de tráfego benigno e malicioso, ajustando seus pesos de forma adaptativa. Essa abordagem superou a detecção baseada em assinaturas, alcançando uma precisão de 98%.

Em um estudo realizado por Tait et al. (2021), diversas técnicas de AM foram comparadas usando conjuntos de dados conhecidos como UNSW-NB15 e CICIDS2017. O estudo mostrou que, para diferenciar entre atividades normais e suspeitas, a Floresta Aleatória teve um desempenho excelente. Para identificar tipos específicos de ataques, o *k-Nearest Neighbor* (k-NN) se destacou, atingindo uma precisão de 99,83%.

Uma subárea do aprendizado de máquina é o aprendizado profundo, do inglês “Deep Learning (DL)”, que envolve o uso de várias camadas de redes neurais para processar informações complexas. Estudos têm mostrado grandes avanços na precisão dos IDS. Segundo Macas et al. (2022), as redes neurais profundas (DNN[11]s) são capazes de identificar padrões complexos de ataques com alta precisão, mas requerem mais recursos de computação para serem treinadas.

Pesquisas usando uma combinação de conjuntos de dados (como CIC-IDS 2017, NSL-KDD, entre outros) demonstraram que DNNs são mais eficazes que métodos tradicionais de AM. Conforme estudo citado por Alboni et al. (2022), uma rede neural profunda foi usada para classificar dados de rede, alcançando uma precisão de 99,96%. A pesquisa destacou a importância dos dados usados e a

11 Tratam-se de redes neurais artificiais que contém uma pilha profunda de camadas ocultas(Géron,2021)

eficácia dos Autoencoders[12], uma técnica que ajuda a detectar atividades suspeitas sem precisar de muitos dados de treinamento.

Outro estudo citado por Alboni et al. (2022) investigou o uso de redes neurais profundas para detectar atividades anormais em redes de computadores. Com base no conjunto de dados NSL-KDD, a pesquisa concluiu que técnicas de aprendizado profundo, como RNN (Rede Neural Recorrente), CNN (Rede Neural Convolucional) e Autoencoders, são mais precisas para detecção de anomalias em sistemas de segurança cibernética do que as técnicas tradicionais de AM.

Hoje em dia, com o aumento do volume de dados, os modelos de AM convencionais estão se tornando menos eficazes, o que leva os pesquisadores a adotarem técnicas de DL, como CNNs. Esses modelos são capazes de aprender e extrair informações úteis dos dados para melhorar a detecção de ameaças, inclusive para ataques do dia zero. No entanto, eles exigem recursos computacionais poderosos e mais tempo para serem treinados, geralmente usando GPUs e plataformas de computação em nuvem.(Vanin, 2022).

A implementação de IDS para sistemas de Controle Industrial (ICS) é um grande desafio. Os ataques cibernéticos aos ICSs representam desafios significativos devido às arquiteturas únicas desses sistemas, que são compostos por hardware de Controle de Supervisão e Aquisição de Dados (SCADA) e software que permite o controle humano das máquinas. Um exemplo notável é o ataque Stuxnet, considerado a primeira arma de guerra cibernética, cujo alvo principal foi provavelmente o programa nuclear iraniano. Os ataques a ICSs podem ser patrocinados pelo Estado, lançados por concorrentes, invasores internos com objetivos maliciosos. Estes ataques podem ter consequências graves, como cortes de energia em larga escala, liberação de substâncias tóxicas e explosões, afetando a saúde pública, segurança nacional e economia.

Portanto, é crucial desenvolver IDS robustos e específicos para ICSs, capazes de proteger contra ameaças cada vez mais sofisticadas. Esses sistemas devem ser seguros, confiáveis e adaptáveis para garantir a segurança e eficiência dos processos industriais.

12 Vide Glossário, pág. 45.

Em análise aos estudos apresentados na tabela A tabela, disponibilizada no Apêndice A, percebe-se que tanto o aprendizado de máquina quanto o aprendizado profundo têm contribuído significativamente para o avanço dos sistemas de detecção de intrusão. Enquanto o AM continua sendo eficaz em várias aplicações, o DL se destaca pela capacidade de lidar com dados complexos e grandes volumes de informações, proporcionando alta precisão na detecção de anomalias. A evolução e o sucesso dessas técnicas dependem, em grande parte, do desenvolvimento de novos algoritmos e da capacidade de computação disponível. Assim, conforme os estudos apresentados, pode-se perceber que a tendência é que, com o aumento dos recursos tecnológicos, o aprendizado profundo se torne cada vez mais predominante na área de segurança cibernética.

4.2.1 Desafios e limitações da aplicabilidade de AM e DL em IDS

A implementação de IDS utilizando aprendizado de máquina e aprendizado profundo apresenta uma série de desafios e limitações, conforme serão evidenciados por Alboni et al (2022) e por Macas et al (2022) a seguir.

Segundo Alboni (2022), as principais diferenças entre aprendizado de máquina e aprendizado profundo são:

- Dependência de dados: Indica o volume de dados. Para o aprendizado de máquina, o desempenho é melhor quando o conjunto de dados é limitado; já o aprendizado profundo tem melhor desempenho com um grande número de dados.

- Processamento de recursos: No aprendizado de máquina, os recursos (características dos dados) a serem utilizados no modelo devem ser indicados pelo especialista; no aprendizado profundo, a representação dos recursos são identificadas automaticamente por meio do uso de algoritmos do aprendizado profundo.

- Interpretabilidade: Refere-se à capacidade de o resultado de um modelo ser entendido por humanos. Um modelo interpretável pode ser entendido sem necessidade de procedimentos extras. Alguns algoritmos de aprendizado de máquina são interpretáveis, no entanto, no aprendizado profundo, é difícil especificar como os neurônios devem ser modelados e como as camadas devem interagir.

- Solução de problemas: No aprendizado de máquina, o problema é dividido em subproblemas, sendo cada um resolvido de forma independente, e então a resposta final é obtida, já no aprendizado profundo o problema é resolvido de forma completa.

Macas et al. (2022) apontam vantagens estratégicas específicas do aprendizado profundo em cibersegurança, que incluem:

Simplicidade**:** O aprendizado profundo simplifica a criação manual de características, substituindo pipelines complexas e frágeis por modelos treináveis de ponta a ponta, reduzindo significativamente o trabalho necessário.

Escalabilidade**:** Os algoritmos de aprendizado profundo não requerem o armazenamento de todos os dados na memória e podem melhorar seu desempenho com grandes volumes de dados, oferecendo melhor escalabilidade em comparação com os métodos tradicionais de aprendizado de máquina.

Reutilização: Modelos de aprendizado profundo podem ser treinados com dados adicionais sem precisar recomeçar do zero e são adequados para treinamento contínuo. Além disso, podem ser reaproveitados através do aprendizado por transferência, permitindo a reinvestimento do trabalho anterior em modelos mais sofisticados e robustos.

Diante das diferenças entre AM e DL trazidas por Alboni et al(2022) e as vantagens apresentadas por Macas et al. (2022), pode-se concluir que, apesar das vantagens do aprendizado profundo, como simplicidade, escalabilidade e reutilização, ambos os métodos possuem desafios e limitações. A escolha entre aprendizado de máquina e aprendizado profundo deve considerar as especificidades do ambiente de aplicação, os recursos disponíveis e as necessidades de interpretabilidade e precisão do IDS.

4.3 Conclusão parcial

Neste capítulo a atual aplicabilidade do aprendizado de máquina e do aprendizado profundo na detecção de intrusões foi explorada, com ênfase na aplicação das técnicas e nos resultados já alcançados. Também foram analisados os principais conjuntos de dados utilizados na construção de Sistemas de Detecção

de Intrusão (IDS). Estudos indicam que a escolha do conjunto de dados é de grande importância para a implementação dos IDS. Observou-se que sistemas IDS construídos com conjuntos de dados mais recentes apresentaram menor precisão em comparação com aqueles baseados em dados mais antigos, nos quais a precisão era excelente. Outro ponto relevante em relação aos conjuntos de dados é a sua ineficiência em detectar classes específicas de ataques, o que ocorre devido ao desequilíbrio entre as classes.

Além disso, foram examinados os principais desafios e limitações enfrentados na aplicação das técnicas de AM na detecção de intrusões. Constatou-se que as técnicas de aprendizado profundo apresentam maior precisão na detecção de anomalias em sistemas de segurança cibernética em comparação com as técnicas tradicionais de AM, além de serem mais eficazes na aprendizagem e extração de informações úteis para a melhoria na detecção de ameaças, inclusive para ataques de dia zero.

5 CONCLUSÃO

Ao longo desta monografia, exploramos a aplicação de técnicas de AM e aprendizado profundo no contexto dos IDS, destacando as vantagens, desafios e limitações dessas abordagens. Os resultados apresentados refletem o estado da arte na detecção de anomalias e indicam direções promissoras para o desenvolvimento futuro dessas tecnologias. A pesquisa objetivou responder a questão de como a Inteligência Artificial tem sido usada para a defesa cibernética. Como o tema pode ser muito amplo nos limitamos nos Sistemas de detecção de intrusão, os quais são essenciais a esse objetivo.

No segundo capítulo, que estabeleceu a base teórica para a compreensão dos capítulos subsequentes, foi apresentado o conceito de IDS, ressaltando sua importância na segurança cibernética contemporânea. Foram analisados os diferentes tipos de IDS, com ênfase especial nos sistemas baseados em assinaturas e anomalias. Ademais, também foram abordadas as técnicas de AM aplicadas aos IDS, sendo destacadas as abordagens baseadas em classificadores, como SVM e

RF, que, apesar de serem eficazes, enfrentam limitações na detecção de ataques mais sofisticados.

No terceiro capítulo, foram analisados casos de ataques cibernéticos, com ênfase naqueles voltados para objetivos militares. Destacam-se os ataques DDoS ocorridos em 2007, que paralisaram servidores essenciais na Estônia, e a ofensiva militar e cibernética coordenada que atingiu a infraestrutura da Geórgia, também utilizando ataques DDoS. Esses exemplos demonstram a vulnerabilidade de infraestruturas críticas e ilustram como ofensivas dessa natureza poderiam ter sido evitadas ou, ao menos, mitigadas por meio da adoção de IDS robustos. A implementação de IDS baseados em técnicas de AM é fundamental, pois esses sistemas, quando devidamente configurados, são capazes de identificar padrões anômalos de tráfego e, assim, detectar e mitigar ataques DDoS em tempo real, evitando prejuízos graves como os observados na Estônia e na Geórgia.

Com o objetivo de identificar os estudos recentes sobre a aplicação de AM em IDS, o quarto capítulo focou na aplicabilidade das técnicas de AM e DL em IDS nos últimos cinco anos. Com base nos estudos revisados neste capítulo, é evidente que as técnicas de AM oferecem um caminho promissor para a detecção de intrusões em redes, com potencial de aprimorar significativamente a precisão e a eficiência dos Sistemas de IDS. Foi realizada uma comparação detalhada entre essas abordagens, levando em consideração os resultados obtidos em diferentes estudos. Constatou-se que, embora o DL ofereça vantagens significativas, as técnicas tradicionais de AM ainda mantêm sua relevância, especialmente em contextos onde há limitações de recursos computacionais. As técnicas de DL, por sua vez, demonstraram superioridade em termos de precisão e capacidade de generalização.

As técnicas supervisionadas, como SVM e ANN, já demonstraram sucesso expressivo na detecção de intrusões. Entretanto, as abordagens de aprendizado profundo, como CNN e RNN, destacam-se por oferecer maior precisão e melhor capacidade de detectar padrões complexos. Ainda assim, a implementação dessas tecnologias enfrenta desafios consideráveis, como a necessidade de dados de alta qualidade, a atualização contínua dos modelos e a demanda de recursos

computacionais, fatores que precisam ser superados para a ampla adoção dessas tecnologias.

No contexto da defesa cibernética, a adoção de IDS baseados em AM pode desempenhar um papel crucial na proteção dos ativos críticos de TIC. A evolução contínua das ameaças cibernéticas exige que essas soluções sejam continuamente aprimoradas, sobretudo no que tange às infraestruturas críticas, como os ICS. O ataque do Stuxnet às instalações nucleares do Irã exemplifica a vulnerabilidade dessas infraestruturas a *malwares* avançados, destacando a importância de IDS robustos e específicos para ICS.

Ademais, questões como a alta taxa de falsos positivos e a detecção de ataques de dia zero continuam sendo desafios que demandam pesquisas e desenvolvimentos adicionais. Estudos futuros devem focar no uso de AM e DL para criar IDS capazes de identificar e mitigar ameaças complexas em tempo real. Na MB, essa abordagem seria fundamental para fortalecer a capacidade de defesa cibernética de sua infraestrutura, garantindo a proteção dos ativos críticos de TIC da MB, com especial atenção especial às forças navais em operação. Além disso, a proteção eficiente dessas infraestruturas contribui para aumentar a resiliência das redes e sistemas críticos do país. Nações que priorizam o fortalecimento de suas defesas cibernéticas estarão mais preparadas para enfrentar esses desafios, assegurando a proteção tanto das redes atuais quanto de futuras ameaças, e, consequentemente, reforçando a posição estratégica do Brasil no cenário global de cibersegurança.

REFERÊNCIAS

BRASIL. **Política Nacional de Defesa. Estratégia Nacional de Defesa**. 79p. 2020a. Disponível em: https://www.gov.br/defesa/pt-br/arquivos/estado_e_defesa/pnd_end_congresso_.pdf. Acesso em: 20 mar 2024.

BRASIL. **Plano Estratégico da Marinha**. Brasília. Marinha do Brasil. 2020b (PEM2020). Disponível em: https://www.marinha.mil.br/sites/all/modules/pub_pem_2040/book.html . Acesso em: 15 ago 2024

BRASIL. Estado-Maior da Armada. **EMA-323**: Política Naval. Marinha do Brasil.2019. Disponível em: https://www.marinha.mil.br/sites/all/modules/politica_naval/book.html. Acesso em: 15 ago 2024.

BRASIL. Estado-Maior da Armada. **EMA-419:** Doutrina Cibernética da Marinha. Brasília, DF, 2021a. Disponível em: http://www.ema.mb/docs/publicacoes/EMA-419-ED2021.aao
Acesso em: 15 ago 2024.

BRASIL. **Portaria GSI N.º 93**, de 18 de outubro de 2021. Aprova o glossário de segurança da informação. Diário Oficial da República Federativa do Brasil. Brasília, DF, 19 dez. 2021b. Ed. 197, Seção 1, pt. 36.

BRASIL. Estado-Maior das Forças Armadas. **MD31-M-07**: Doutrina militar de defesa cibernética. 2. ed. Brasília, DF: Ministério da Defesa, 2023. Disponível em: https://www.gov.br/defesa/pt-br/assuntos/estado-maior-conjunto-das-forcas-armadas/doutrina-militar/publicacoes-1/publicacoes/MD31M07DoutrinaMilitardeDefesaCiberntica2Edio2023.pdf. Acesso em: 15 ago. 2024.

BRASIL. Diretoria-Geral do Material da Marinha. **Plano de Direção setorial do Material.** 2024. Disponível em: https://www.dgmm.mb/sites/default/files/PDS2024.pdf Acesso em : 01/09/2024

BREIMAN, L. Machine learning, volume 45, number 1 - springerlink. *Machine Learning*, v. 45, p. 5–32, 10 2001.

FACELI, Katti*;* LORENA, Ana Carolina; GAMA; João; ALMEIDA, Tiago A.; CARVA, André C.P.L.F. **Inteligência Artificial: Uma abordagem em aprendizado de máquina**. 2 ed. Rio de janeiro: LTC, 2021.

Géron, Aurélien. **Mãos A Obra: Aprendizado De Máquina Com Scikit-Learn, Keras & TensorFlow: Conceitos, Ferramentas e Técnicas Para a Construção de Sistemas Inteligentes** (Portuguese Edition). Alta Books.2021. Edição do Kindle.

HALBOUNI, Asmaa; GUNAWAN, Teddy Surya; HABAEBI, Mohamed Hadi; HALBOUNI, Murad; KARTIWI, Mira; AHMAD, Robiah. **Machine Learning and Deep Learning Approaches for CyberSecurity: A Review.** IEEE Xplore. Vol. 10 (2022). Disponível em: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9712274 . Acesso em: 23 março 2024.

ZETTER, Kim. Contagem Regressiva até Zero Day. Editora Brasport, 2014.

TAIT, Kathryn-Ann; KHAN, Jan; ALGAHTANI, Fehaid; Shah, Awais; KHAN, Fadia; REHMAN, Mujeeb; BOULILA, Wadii; AHMAD, Jawad. **Intrusion Detection using Machine Learning Techniques: An Experimental Comparison**. Paper, *International Congress of Advanced Technology and Engineering (ICOTEN)*, Taiz, Yemen, 2021. Disponível em: https://ieeexplore.ieee.org/document/9493543. Acesso em: 24 março 2024.

GUPTA, A. R. B. AND AGRAWAL, J., "**A Comprehensive Survey on Various Machine Learning Methods used for Intrusion Detection System**," *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)*, Gwalior, India, 2020. Disponível em: https://ieeexplore.ieee.org/document/9115764. Acesso em: 24 março 2024.

VANIN, P., NEWE, T., DHIRANI, L. L., O'CONNELL, E., O'SHEA, D., LEE, B., & RAO, M. (2022). **A study of network intrusion detection systems using artificial intelligence/machine learning.** *Applied Sciences*, *12*(22), 11752. Disponível em: https://www.mdpi.com/2076-3417/12/22/11752. Acesso em: 23 abril 2024.

CLARKE, R. A.; KNAKE, R. K.**Guerra cibernética**: a próxima ameaça à segurança e o que fazer a respeito. 1. ed. Rio de Janeiro: Brasport, 2015.*E-book*. Disponível em: https://plataforma.bvirtual.com.br. Acesso em: 24 jun. 2024.

MACAS, Mayara; WU, Chunming; FUERTES, Walte. **A survey on deep learning for cybersecurity: Progress, challenges, and opportunities**. Comput. Netw. 212, C (Jul 2022). Disponível em: https://doi.org/10.1016/j.comnet.2022.109032. Acesso em: 24 março 2024.

Khraisat, Ansam & Gondal, Iqbal & Vamplew, Peter & Kamruzzaman, Joarder. (2019). **Survey of intrusion detection systems: techniques, datasets and challenges.** Cybersecurity. 2. 10.1186/s42400-019-0038-7.

Lamba, Anil & Singh, Satinderjeet & Bhardwaj, Sachin & Dutta, Natasha & Sai, Sivakumar & Muni, Rela. (2015). **Uses of Artificial Intelligent Techniques to Build Accurate Models for Intrusion Detection System**. SSRN Electronic Journal. 10.2139/ssrn.3492675.

Murel, Jacob, Kavlakoglu, Eda. **What is ensembling learning?**IBM. 2024.

Disponível em: https://www.ibm.com/topics/ensemble-learning . Acesso em: 22 agosto 2024.

Vargas, A. C. G., Paes, A., & Vasconcelos, C. N. (2016, October). **Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres**. In Proceedings of the xxix conference on graphics, patterns and images (Vol. 1, No. 4). sn. Disponível em: http://gibis.unifesp.br/sibgrapi16/eproceedings/wuw/7.pdf. Acesso em: 01/09/2024.

Salehinejad, Hojjat; Sankar, Sharan; Barfett, Joseph; Colak, Errol ; Valaee, Shahrokh. 2018. **Recent Advances in Recurrent Neural Networks**. ArXiv :1801.01078. Disponível em:https://arxiv.org/abs/1801.01078. Acesso em: 01/09/2024.

GLOSSÁRIO

Autoencoders - são redes neurais artificiais que aprendem representações densas e compactas dos dados de entrada sem supervisão. Essas codificações possuem dimensionalidade reduzida, facilitando a visualização e a detecção de características. Além de serem úteis na redução de dimensionalidade, também podem ser usados para pré-treinamento de DNN e, em alguns casos, são capazes de gerar novos dados semelhantes aos de treinamento. (Géron, 2021)

Ataque Cibernético - compreende ações para interromper, negar, degradar, corromper ou destruir informações ou sistemas computacionais armazenados em dispositivos e redes computacionais e de comunicações do oponente. (Brasil, 2023)

bot - tipo de *malware* que, além de incluir funcionalidades de *worms* , dispõe de mecanismos de comunicação com o invasor, os quais permitem que o computador infectado seja controlado remotamente. O processo de infecção e propagação do bot é similar ao do *worm* , ou seja, o bot é capaz de se propagar automaticamente, explorando vulnerabilidades existentes em programas instalados em computadores. (Brasil, 2021b)

Bootnet - rede formada por diversos computadores zumbis (infectados com bots). Permite potencializar as ações danosas executadas pelos bots e ser usada em ataques de negação de serviço, esquemas de fraude, envio de spam , entre outros. (Brasil, 2021b)

CNN - Convolutional Neural Network - é uma variação das redes de Perceptrons de Múltiplas Camadas, tendo sido inspirada no processo biológico de processamentos de dados visuais. De maneira semelhante aos processos tradicionais de visão computacional, uma CNN é capaz de aplicar filtros em dados visuais, mantendo a relação de vizinhança entre os pixels da imagem ao longo do processamento da rede (Vargas et. al)

Defesa Cibernética - conjunto de ações ofensivas, defensivas e exploratórias, realizadas no Espaço Cibernético, no contexto de um planejamento nacional de nível estratégico, coordenado e integrado pelo Ministério da Defesa, com as finalidades de proteger os sistemas de informação de interesse da Defesa Nacional, obter dados para a produção de conhecimento de Inteligência e comprometer os sistemas de informação do oponente. (Brasil, 2023)

DoS – Negação de Serviço - bloqueio de acesso devidamente autorizado a um recurso ou a geração de atraso nas operações e funções normais de um sistema, com a resultante perda da disponibilidade aos usuários autorizados. O objetivo do ataque DoS é interromper atividades legítimas de um computador ou de um sistema. Uma forma de provocar o ataque é aproveitando-se de falhas ou de vulnerabilidades presentes na máquina vítima, ou enviar um grande número de mensagens que esgotem algum dos recursos da vítima, como CPU, memória, banda, entre outros. Para isto, é necessária uma única máquina poderosa, com bom processamento e bastante banda disponível, capaz de gerar o número de mensagens suficiente para causar a interrupção do serviço. (Brasil, 2021b)

DdoS – Negação de Serviço Distribuída (*DDoS*) - atividade maliciosa, coordenada e distribuída, em que um conjunto de computadores ou de dispositivos móveis é utilizado para tirar de operação um serviço, um computador ou uma rede conectada à Internet. Embora os ataques do tipo DoS sejam, em geral, perigosos para os serviços de Internet, a forma distribuída é ainda mais perigosa, justamente por se tratar de um ataque feito por várias máquinas, que podem estar espalhadas geograficamente e não terem nenhuma relação entre si, exceto o fato de estarem parcial ou totalmente sob controle do atacante. (Brasil, 2021a)

Ensembling Learning - Aprendizado em conjunto (*ensemble learning*) é uma técnica de aprendizado de máquina que agrega dois ou mais modelos (por exemplo, modelos de regressão, redes neurais) para produzir previsões melhores. Em outras palavras, um modelo em conjunto combina vários modelos individuais para gerar

previsões mais precisas do que um único modelo isolado. (Murel e Kavlakoglu, 2024)

Firewalls - ferramenta para evitar acesso não autorizado, tanto na origem quanto no destino, a uma ou mais redes. Podem ser implementados por meio de hardware ou software , ou por meio de ambos. Cada mensagem que entra ou sai da rede passa pelo *firewall* , que a examina a fim de determinar se atende ou não os critérios de segurança especificados. (Brasil,2021b)

malware - software malicioso, projetado para infiltrar um sistema computacional, com a intenção de roubar dados ou danificar aplicativos ou o sistema operacional. Esse tipo de software costuma entrar em uma rede por meio de diversas atividades aprovadas pela empresa, como e-mail ou sites . Entre os exemplos de *malware* estão os vírus, *worms* , trojans (ou cavalos de Troia), *spyware* , *adware* e *rootkits* (Brasil, 2021b)

Probe – Probing attacks - Ter acesso a todas as informações da rede antes de iniciar um ataque. (Lambda et. al)

R2L – Root to Local attacks - Ao explorar algumas vulnerabilidades da rede, o invasor obtém acesso local enviando pacotes para uma máquina remota. (Lambda et. al)

U2R – User to root attack - Inicialmente, o invasor acessa uma conta de usuário comum e, posteriormente, ganha acesso ao root explorando as vulnerabilidades do sistema. (Lambda et. al)

RNN – (Recurrente Neural Network) - Redes neurais artificiais (ANNs) com conexões recorrentes são chamadas de redes neurais recorrentes (RNNs), que são capazes de modelar dados sequenciais para reconhecimento e previsão de sequências. (Salehinejad et.al, 2018)

Segurança Cibernética - arte de assegurar a existência e a continuidade da sociedade da informação de uma nação, garantindo e protegendo, no Espaço Cibernético, seus ativos de informação e suas infraestruturas críticas. (Brasil,2021b)

APÊNDICE A – Tabela Resumo dos trabalhos analisados.

Autores	**Algoritmos utilizados**	**Melhores resultados**	**Classificação**
Halboni et al., 2022	SVM, RF	SVM teve desempenho superior no treinamento; RF no teste	AM
Halboni et al., 2022	Árvore de Decisão, Naive Bayes, RF	Melhor precisão para ataques DoS com RF	AM
Halboni et al., 2022	Redes Neurais Artificiais (RNAs)	98% de precisão na detecção de tráfego malicioso	AM
Tait et al. (2021)	ANN, K-NN, LR,RF	K-NN obteve 99,83% de precisão na classificação multiclasse	AM
Halboni et al (2022)	Deep Neural Network (DNN)	Alta precisão e baixa taxa de alarmes falsos	DL
Halboni et al (2022)	Deep Neural Network (DNN), Autoencoder	Precisão de 99,96% com UNB-ISCX 2012 e CIC-IDS 2017	DL
Halboni et al (2022)	RNN, CNN, Autoencoder	RNN teve melhor desempenho, seguido por CNN e Autoencoder	DL